

World Conference on Educational Sciences 2009

Spanish students and teachers' preferences towards computer-based and paper-and-pencil tests at universities

De-Siqueira, Jose Macario^{a*}, Peris-Fajarnes, Guillermo^a, Gimenez, Fernando^a, Magal-Royo^a, Teresa^a

^a*Universidad Politécnica de Valencia, Valencia 46022, Spain*

Received October 22, 2008; revised December 10, 2008; accepted January 02, 2009

Abstract

This study was conducted to identify and analyze some significant features that influence students and teachers about computer-based tests (CBT) and paper-and-pencil tests (P&P) at the context of the PAULEX Project. In order to do that, a large experiment has been developed at the Universidad Politécnica de Valencia (Polytechnic University of Valencia), Spain, in which several students and professors have answered a validated questionnaire about their usage of technology, feelings and experiences. They also compared their preferences after doing two similar basic tests, CBT and P&P.

© 2009 Elsevier Ltd. All rights reserved

Keywords: Assessment; computer-based tests; paper-based tests; test performance.

1. Introduction

The use of Information and Communication Technologies, specially increased by mobile technology and the Internet, is changing many aspects of the everyday life. People are normally connected and interacting with computers, mobile phones, palmtops, laptops, videogames and global positioning system (GPS) devices, interconnected by a variety of communication technologies such as wireless fidelity (WiFi), Bluetooth, dial-up services and virtual private networks. As these technologies are becoming familiar to many people, they can be used in several educational contexts, including the student assessment through computer-based tests (CBT) and others.

There are many studies dealing with the use of computers for testing and assessment, analyzing aspects such as Computer Experience and Computer Anxiety (e.g. Smith & Caputi, 2007; Chua, Chen, & Wong, 1999; Mahar, Henderson, & Deane, 1997), some of them comparing differences and equivalences between CBTs and paper-and-pencil (P&P) tests (e.g. McDonald, 2002; Norris, Pauli, & Bray, 2007; Russel & Haney, 1997). Nevertheless few articles in major journals have focused on analyzing the relationship between the usage of technology in the lifelong learning and their preferences and results comparing CBT and P&P.

This study was conducted to identify and analyze some significant features that influence students and teachers about CBT and P&P in the context of the PAULEX-Universitas Project, led by the Department of Applied Linguistics of the Universidad Politécnica de Valencia (Polytechnic University of Valencia), Spain, and funded by

E-mail address: jodesi@doctor.upv.es

the Spanish Ministry of Science and Innovation. The principal aim of the PAULEX-Universitas project is the development of a computer tool for the realization of the University Entrance Language Exams in Spain (García Laborda, 1999). In order to do that, a large experiment has been developed at this University, in which 83 undergraduate students, 19 postgraduate students and 22 professors have answered to a validated questionnaire about the way in which they use technology, their feelings and experiences. They also compared their preferences after doing two similar basic tests, CBT and P&P. Some statistically significant issues were found, specially related to the computer experience as the principal factor on determining participants' preferences. Another interesting result is that a substantial number of people consider the CBT as more appropriate and they effectively have better results in spite of feeling insecure while doing the test via Internet. They also consider the P&P safer.

2. Method

The study was conducted at the Polytechnic University of Valencia, an important public University in Spain, as one of the first experiments framed in the PAULEX Project. As a qualitative method, several experts on Computer-Assisted Language Learning and Testing (CALT) and the responsible person for the University Entrance English Test in Valencia have been previously interviewed in order to compile information about their experiences and beliefs concerning the significant features which could influence the results if the University Entrance Exams were computerized. Then, those findings connected to the students' and teachers' preferences towards CBT and P&P have been considered in order to prepare the bases of this experiment. A validated questionnaire had been created for quantitative analyses providing statistically significant findings and two similar tests had been developed to simulate the context of use of a CBT and a P&P.

All participants were volunteers involved in language learning from the same University. The 124 participants who completed the tests and the questionnaire correctly were divided into three groups: 83 were undergraduate students (mean age 22), 19 postgraduate students (mean age 29) and 22 professors (mean age 43). From the 83 undergraduate students group, 45 were Erasmus students from other European countries (18 males and 27 females) and 40 were Spanish students (24 males and 16 females). 13 out of the 19 postgraduate students (7 males and 12 females) came from the Department of Applied Linguistics and all professors were language teachers (4 males and 18 females). Some excluded participants did not complete the whole questionnaire and/or failed doing the tests (5 undergraduate students and 3 professors).

The tests were developed taking into account pedagogical parameters and psychometric tasks concerning language testing in order to make them similar and valid for the assessment of Spanish language students. They were developed by a Spanish professor, reviewed by other four and accepted by all language professors who participated in this experiment. They were compounded by a current and real text from the Internet of about 500 words, a multiple-choice section with 10 questions and a writing section with four questions. The principal objective was simulating the task of doing the tests, prioritizing aspects such as the time and preferences of format instead of the students' answers, always bearing in mind the questionnaire. The tests were not very long (about 10 minutes each) and the level of Spanish was intermediate. Both tests were made in both formats, CBT and P&P. As the contents of the tests could not affect the comparison of formats for the questionnaire, the half of the participants did the first test as P&P and the second as CBT, and the rest of participants did the same first test as CBT and the second as P&P. The use of different multimedia and interactive elements was avoided in order to make both formats as similar as possible. Despite of all, excluding some of the possibilities offered by technology for CBT and making it similar to a P&P limited the results of the experiment by using just the first of the three generations of testing systems argued by Bennett (1998). He defines that the first generation takes limited advantage of technology, making it very similar to the P&P ones, the second includes new possibilities such as multimedia elements and the automatic item generation, and the last one consists of continuous assessment during the learning process. Currently there are Learning Management Systems (LMS), as the Ingenio System developed by the Camille Research Group and used for the CALL@C&S Project, co-funded by the Lingua Project of the European Commission (Gimeno, 2005), which are consolidating their tools related to this third generation of testing systems.

The questionnaire, which lasted about 5 minutes and which consisted of 25 multiple-choice questions, was developed with then intention of not being exhaustive, as the experiment wouldn't take excessive time to be completed. The questions were divided into three sections:

- 1- Personal Characteristics
- 2- Computer Experiences and Anxiety
- 3- Comparing beliefs, preferences and feelings about CBT and P&P.

The first section, with 4 questions, compiles personal characteristics to classify participants into different groups in order to identify parameters which could influence the questionnaire results. The second section, with 6 questions, compiles information about Computer Anxiety and Computer Experience taking into account several studies and validated scales (e.g. Gaudron & Vignoli, 2002; Smith & Caputi, 2007; Mahar, Hederson, & Deane, 1997). The last section, with 15 questions, recognizes information about preferences and feelings from participants when comparing CBT and P&P. All questions of the second section and 10 of the third section use Linkert scales.

Each participant, after reading brief instructions and asking the responsible for procedures, completed first the CBT, then the P&P and finally the questionnaire. While doing the CBT, the system was automatically saving on the server all the important information from the participants, including answers and the time spent on each section of the test. On P&P, students had to mark by themselves the time spent on each section. Once completed the questionnaire, a full report was prepared recompiling information from the CBT, the P&P and the questionnaire. All the language test answers were included in this report in order to confirm that participants had paid enough attention and had answered all questions satisfactorily. These reports, including all the test answers and specially the free writing section, have been used for other studies about students' results of CBT and P&P. All information from questionnaires and tests was ported into a statistical analysis package program (SPSS) for a significant statistical analysis (level of 0.05).

3. Results

Several significant features that influence participants' preferences towards CBT and P&P were found, including personal characteristics and experiences and the pedagogic analysis of the test results. The main aim of this experiment was the identification of these features within the frame of the PAULEX Project. Nevertheless, the principal objective of this article has been the compilation of the questionnaire results.

The first important finding was that, using simple factorial ANOVAs analysis, some statistically significant differences and other unimportant ones were identified. All the statistical analyses were reconsidered by classifying participants into men and women in order to look for sex differences, but no significant differences were found. With regard to the participants' experiences and usage of technology, the age was a significant factor because the younger the participants were, the more experience they had. Similar results were found when classifying the participants into undergraduate and postgraduate students and professors.

The undergraduate students, taking into account differences of age and experience with computers, tended to prefer the CBT instead of P&P, considering the CBT as more appropriate (91%) and preferable (96%) at universities and they also previewed that the CBT would overcome the P&P in less than four years at the Spanish universities. They defended that students are already prepared for the CBT, but teachers (68%), materials (78%) and computer tools (86%) are not. The professors' and postgraduate students' answers to the questionnaires were much more diverse, including difference of age, computer experiences and believes when comparing CBT and P&P. Only a half of professors and 78% of postgraduate students considered the CBT as more appropriate and preferable at universities. As one unique group, 70% of professors and postgraduate students considered that students are prepared to use CBT, only 46% considered to be prepared to use CBT in their classes and all of them defended that there are not available and adequate materials to be used. Only the five professors who are using some tools of the *InGenio* System to create CBT for their students support that there are acceptable computer tools to be used, in spite of thinking that some improvements are required. Another interesting result is that, despite these preferences considering CBT as preferable and more appropriate, and in spite of better results in the written section, most of the participants (69% of undergraduate students and 82% of others) feel insecure when doing the test via Internet and also consider the P&P safer.

Concerning the time, important differences were found when comparing the CBT and the P&P results. In regard with readability, only four students out of all the participants preferred reading the text on a computer screen, and there are not significant differences in relation to the time spent on reading. No important time differences were found in the multiple-choice section, and 95% of participants were indifferent about doing this section as CBT or P&P. Most of the important differences were recognized in the writing section. Some important deviations were

identified in the case of some participants who had more difficulties when writing on computers. Undergraduate students saved 30% of time, postgraduate students saved 27% of time and professors saved 14% of time. In addition, 96% of the undergraduate students, 94% of the postgraduate students and 86% of the professors preferred doing this section as CBT. This preference, in comparison with the frequency of using computers for writing, had been proved in SPSS as a statistically significant feature ($P < 0.05$). All participants who write more than 100 words daily or weekly preferred the CBT writing format, while 92% of those who seldom or sometimes use computers for writing chose the P&P format.

4. Conclusion

This experiment reached its main goal, which was to provide significant findings in relation to the preferences when comparing the CBT and the P&P test formats. The procedures, initially based on previous studies, had been discussed and approved by several experts who are part of the PAULEX Project and permitted the recognition of vast information by the participants. This paper aimed to find some special issues about the current state of students' and teachers' experiences and preferences concerning the use of CBT for Language Testing at the Polytechnic University of Valencia. It could identify some similarities and differences by classifying participants into groups; for example, professors were less familiarized with CBT than students and no important sex differences were found. Another finding was that most participants prefer writing on CBT than on P&P, which permits them to save a considerable amount of time, and they also consider as more appropriate the use of CBT at university despite they feel insecure when using the Internet and are aware of their computer fails. The findings of this paper have important implications for the academic community when considering the impact of usage of technology on high-stakes educational assessment in Spain, providing initial evidences for latter empirical studies and especially concerning the CBT via Internet as the probably future substitute of P&P at universities.

5. Acknowledgements

Acknowledgements are due to the Spanish Ministry of Science and Innovation for funding the PAULEX Project, to all participants of this experiment for their generous collaboration and to the Camille Research Group sponsors for cooperating.

References

- Bennett, D. (1994). *Reinventing assessment: speculations on the future of large-scale educational testing*. Princeton, NJ: Educational Testing Service, Policy Information Center.
- Chua, S.L., Chen, D.T., & Wong A.F. (1999). Computer anxiety and its correlates: a meta-analysis. *Computers in human Behaviour*, 15, 609-623.
- García Laborda, J. (2006). "¿Qué pueden aportar las nuevas tecnologías al examen de Selectividad de inglés? Un análisis prospectivo". *Revista de Ciencia de la Educación*, 206, 151-166.
- Gaudron, J.P., Vignoli, E. (2002). Assessing computer anxiety with the interaction model of anxiety: development and validation of the computer anxiety trait subscale. *Computers in Human Behavior*, 18, 315-325.
- Gimeno, A. (2005). New challenges in developing an online CALL authoring shell, content manager and courseware: the INGENIO model, *EUROCALL Newsletter*, 7.
- Kenyon, D. M., & Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning & Technology*, 2, 60-83.
- Kveton, P., Jelinek, M., Voboril, D., & Klimusova, H. (2007). Computer-based tests: the impact of test design and problem of equivalency. *Computers in human Behaviour*, 23, 32-51.
- Mahar, D., Henderson, R., & Deane, F. (1997). The effects of computer anxiety, state anxiety, and computer experience on users' performance of computer based tasks. *Personality and individual differences*, 22, 683-692.
- McDonald, A.S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*, 39, 299-312.
- Norris, J.T., Pauli, R., & Bray, D.E. (2007). Mood change and computer anxiety: A comparison between computerised and paper measures of negative affect. *Computers in human Behaviour*, 23, 2875-2887.
- Russel, M., & Haney, W. (1997). Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil. *Education Policy Analysis Archives*. Available from: <<http://epaa.asu.edu/epaa/v5n3.html>> (10 January 2008).
- Smith, B., & Caputi, P. (2007). Cognitive interference model of computer anxiety: Implications for computer-based assessment. *Computers in human Behaviour*, 23, 1481-1498.